

2009.3.16 セマンティックWebコンファレンス

# セキュアな企業内情報利用に向けて ～情報漏洩防止のための セマンティック技術～

株式会社富士通研究所  
津田 宏

Copyright 2009 株式会社富士通研究所

## 概要

FUJITSU

2007年に報道された個人情報漏えいインシデントの件数は864件。損害賠償総額は2兆円を超えており(NPO日本ネットワークセキュリティ協会調べ)、PC/USB持ち出しや、メールなどネット経由による情報漏洩対策は企業にとって今解決すべき重要な課題と言える。

一方、セマンティックWebが目指しているように、社内外での情報の統合される世界では、情報の活用が容易になる反面、情報漏洩も複雑化しその対策や情報管理も変わっていきと考えられる。

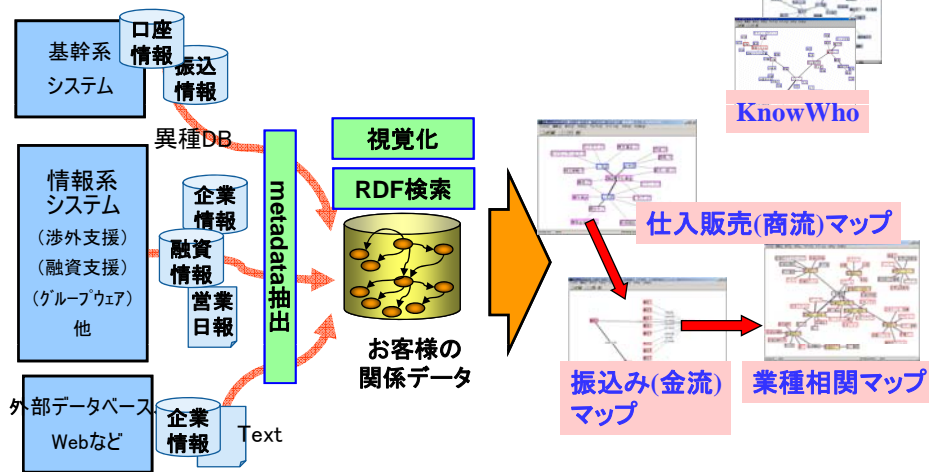
多様な出口からの情報漏洩に向けた根本的な対策として、PCやネットワークなどのエンドポイントを守るだけでなく、情報そのものを守る「情報セントリックセキュリティ」が提案されている。この実現にはセキュリティ技術だけでなく、情報検索や、自然言語処理などのセマンティック技術も必要となる。

本発表では、メールからの情報漏洩防止に対して、宛先ミスなどの過失から機密文書内容の誤送信まで、セマンティック技術を含むレベルに応じた取組みをデモを交えて紹介する。

参考: メールからの情報漏洩対策技術を開発 ～宛先ミスから機密情報流出までレベルに応じた対策が可能に～, 2009.3.13 富士通プレスリリース

# 1. はじめに

## ■ RDFによる社内情報の統合・見える化 (-2007)



2008.5.7 プレスリリース「滋賀銀行様においてビジネス情報ナビゲーションシステムが稼働  
～地域企業のビジネス相関図を見える化することで、地域密着型の提案を実現～」

# 統合の次に来るもの

## ■ 情報統合は良いことばかりか

### ■ プライバシー侵害

- ある委員会の名簿がネットに流出  
→ 論文や学会などの公開情報から、組織＋人名で、e-mail、画像などを加えてマッシュアップ → ……

### ■ 情報漏洩

- 個々の機密情報はタグをつけるなりして、社外への流出ブロックできるだろう。
- だが、二次情報、部分的に組み合わせた文書を作っても大丈夫か。今後、SaaS・クラウドで社内情報が物理的には社外にあるのが当然の世の中でも大丈夫だろうか。

## 2. メールからの情報漏洩対策

ビジネスユーザーの  
66.2%がメール誤送信の  
経験  
誤送信の影響: **お詫び(4.1%),**  
**始末書(1.8%),取引に影響**  
**(1.2%),解雇(0.3%)**

HDE「メール誤送信」に関する実態調査  
2008.4.23

2008年上半期  
漏えい人数: 170万人  
インシデント件数: 683件 (増加傾向)  
漏えい原因: **誤操作(36.8%),**管理ミス  
(19.2%), 盗難(15.7%),紛失(14.5%)  
経路: 紙(55.2%), USB(10.2%), Web(10%),  
PC(8.8%), **mail(8.5%)**

NPO日本ネットワークセキュリティ協会  
「2008年上半期情報セキュリティインシデ  
ントに関する調査報告書」2009.2.20

### ■ メール誤送信対策のニーズ

- メールによる個人情報漏洩の対策
- 他社向けメール取り違え等による企業としての信用低下防止

ただし、メール誤送信に対して何をすれば何が解決するのか、はっきりしない

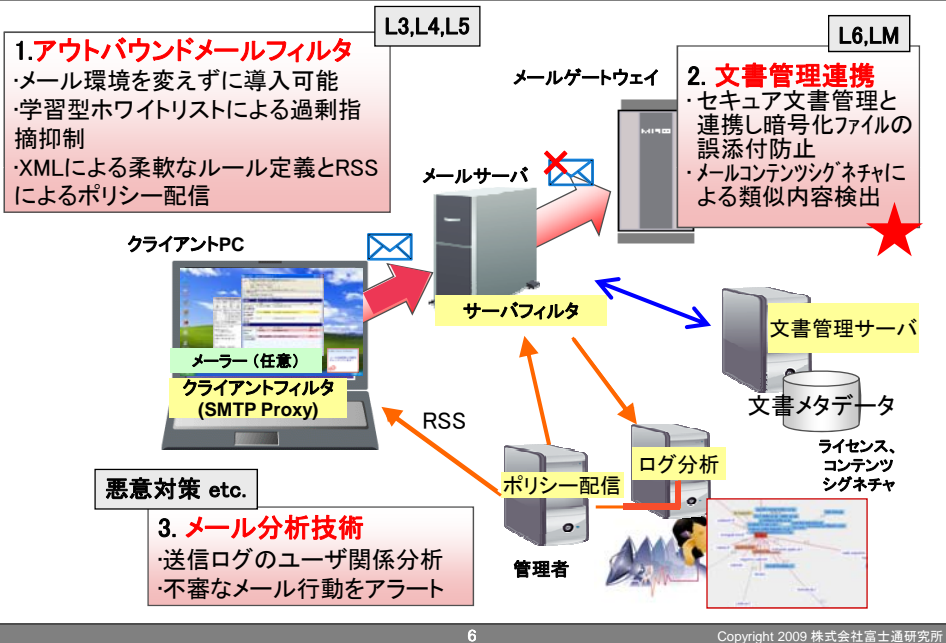
## メール誤送信の対策レベル

レベル	対策内容	インシデントなど
フェール セーフ	L1 添付ファイル暗号化	
	L2 添付ファイルはサーバ経由 で閲覧させる	
汎用的 対策	L3 確認アラートUI: 送信前に再 度気づかせる	(eg) アドレス帳で隣を選択、タイプミ ス、社外のCclに気付かず返信
	L4 ルールに抵触するメールを 禁止	(eg) アドレス帳選択ミスで社外を含 む200名以上に送信
業務 特化 の 対策	L5 プロジェクト・業務ルールに 従ったメールのみ許可	(eg) A社向けファイルをB社に送付
	L6 <b>セキュア文書管理と連携</b>	(eg) 社外秘ファイルを社外に誤添付
	L7 業務を見直し、上司を含めて 内容を承認で確認	(eg) 添付ファイルの内容上の誤り
悪意対策	<b>証跡ログ分析により監視</b>	(eg) 退職前に特定の社外宛メール増

※L1~L7: 下のレベルまで対応するとカバー範囲は広いが、導入の人的コストは高い

今回の注目部分

### 3.メール情報漏洩対策:アプローチ

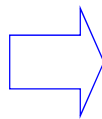


### コンテンツの流出ブロック

#### ■ 誤添付を防止したいコンテンツ

##### ■ 個人情報

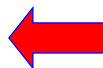
- ・住所、氏名
- ・従業員番号
- ・社会保障番号
- ・...



情報抽出、固有名抽出など、自然言語処理の技術が色々提案されている

##### ■ 機密情報

- ・社外秘、関係者外秘
- ・他社機密情報



今回の対象

#### 【課題】

- ・機密情報をどう指定し、登録するか  
 (「関係者外秘」の文字列有無では単純すぎ。決算資料のようにある時期以降は公開文書になる場合も)
- ・どこまでの流出をブロックするか(機密文書そのもの、一部、改訂)

# メール本文からの情報漏洩防止

## ■ テキストレベルでの文書類似性を検出できるメタデータ

解決したいシーン: お客様あてのプロジェクト報告書の中に、未発表表と知らずに決算資料の一部をコピー、編集して挿入

### コンテンツシグネチャ demo1

元テキスト(機密文書コンテンツなど):

経常利益は14.5億円の損失と、前年同期比では88.2億円の悪化となりました。営業外損益は11.0億円の悪化となりました。欧州の合併会社に係る持分法損益がパソコン、PCサーバの競争激化により悪化したほか、当第3四半期そのる数回の高値の取引に係る影響が拡大しました。グローバル株式会社などの株式の売却益2.9億円を特別利益に計上しました。一方、閉鎖を決定したHDD用ヘッドの製造ラインや電子部品事業などに係る減損損失7.4億円及び米国 Spansion, Inc. など時価が著しく下落した上場株式に係る評価損6.4億円を特別損失に計上しました。なお、上場株式に係る評価損については金額的重要性が高まったため、第2四半期連結結果計期間においては営業外費用に計上していた2.3億円を含めて特別損失に計上しています。

当第3四半期累計(前)半期純利益は36.1億円の損失と、前年同期比32.3億円の悪化

検査対象テキストの一部を切り取って多少編集したようなもの、など

報告年月: 2009年2月度

- 1. AAAシステムの開発状況  
予定通り進捗。3月1日にはテスト工程に入る予定。
- 2. BBBシステムの開発状況  
予定通り進捗。3月1日より人員を増強し、フェーズ2の開発を開始。
- 3. 関連ビジネスの概要:

当社で閉鎖したHDD用ヘッドの製造ラインや電子部品事業などの減損74億円と時価の下落した上場株式の評価損64億円が特別損失に計上される予定です。上

類似度チェック

### 社外向け報告書

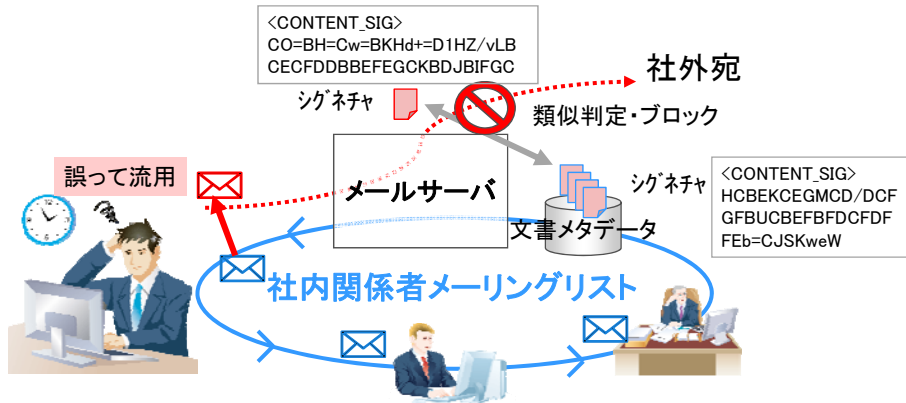
XX建設(山田様) 敬愛に  
お返事になっております。  
Y社受発注の御件です。  
今月のプロジェクト報告をお送りします。よろしくお願いたします。  
プロジェクト名称: ZZZプロジェクト  
報告年月: 2009年2月度  
1. AAAシステムの開発状況  
予定通り進捗。3月8日にはテスト工程に入る予定。  
2. BBBシステムの開発状況  
予定通り進捗。3月1日より人員を増強し、フェーズ2の開発を開始。  
3. 関連ビジネスの概要:  
当社で(2008)閉鎖した(5124)HDD用ヘッドの(7466)減損74(4202)億円と(2008)時価の(1827)下落した(4281)上場株式の(3489)特別損失に(4264)計上される予定です。上(4283)場株式の(3481)評価損は、営業外費用に(4264)計上していた2.3(1202)億円を含めて(3483)特別損失と

未発表資料の一部をコピー&編集して挿入

類似部分を検出・アラート

# 利用シーン

・機密MLIに流れるメールのコンテンツシグネチャをサーバで自動記録  
・コンテンツシグネチャ:シグネチャファイル検索方式を拡張し、単語の位置関係に基づく文書の特徴情報。元文書の任意の部分をコピー、削除、挿入など編集してもシグネチャの類似性で検出可。シグネチャ自体にはキーワード、個人情報含まない



# コンテンツシグネチャ: 原理

## ■ テキスト中の単語間の位置関係を特徴素

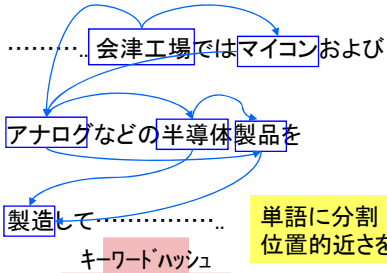
原文

..... 会津工場では  
マイコンおよびアナログ  
などの半導体製品を製  
造して.....

シグネチャにはキー  
ワード、個人情報  
は含まれない

チェック対象

..... 会津は  
マイコンやアナログ半導  
体の製造工場であり、...  
.....



単語に分割  
位置的近さを判定

CO=BH=Cw=BKHd+=D1HZ/vLBCECF  
DDBBEFEGCKBDJBIFGC

シグネチャ(特徴素)

シグネチャ同士を比較することで、  
類似部分を検出(原文不要)

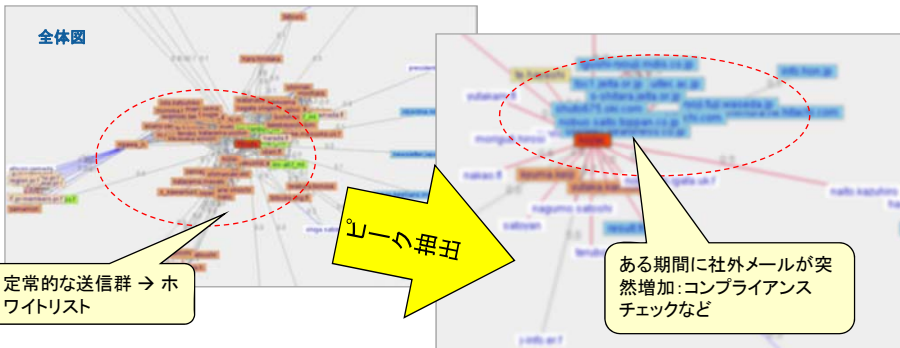
HCBEKCEGMCD/DCFGFBUC  
BEFBFDCFFFEb=CJSKweW

# (参考) ログ分析技術

## 利用シーン

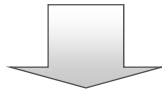
- ・ホワイトリストの定期的な更新(送信チェックの過剰指摘の軽減)
- ・特に最近特定社外とのやりとりが増加などを管理者にアラート

cf. 既存技術: 2008.5.7プレスリリース 滋賀銀行様においてビジネス情報ナビゲーションシステムが稼働~地域企業のビジネス相関図を見える化することで、地域密着型の提案を実現~



### ■ 情報セントリック(中心)セキュリティ

- 従来のセキュリティ: PC対策、USB対策、ファイアウォールなど、エンドポイントの対策や、外からの脅威の対策
- これから: 情報漏洩対策としては、メールやUSB、Webなど様々な経路で中からの脅威への対策が必要。→ エンドポイントではなく、情報そのものを守るアプローチが必要 = 「情報セントリックセキュリティ」
- SaaSやマッシュアップで、情報がさらに自由に組み合わせることのできる時代では更に情報セントリックの考え方が重要に



**セマンティック(Web)技術の期待される領域**